(51) International Patent Classification 5 :		(11) International Publication Number:	WO 91/03042
G10L 9/00	**A1**	(43) International Publication Date:	7 March 1991 (07.03.91)

(54) Title: A METHOD AND AN APPARATUS FOR CLASSIFICATION OF A MIXED SPEECH AND NOISE SIGNAL
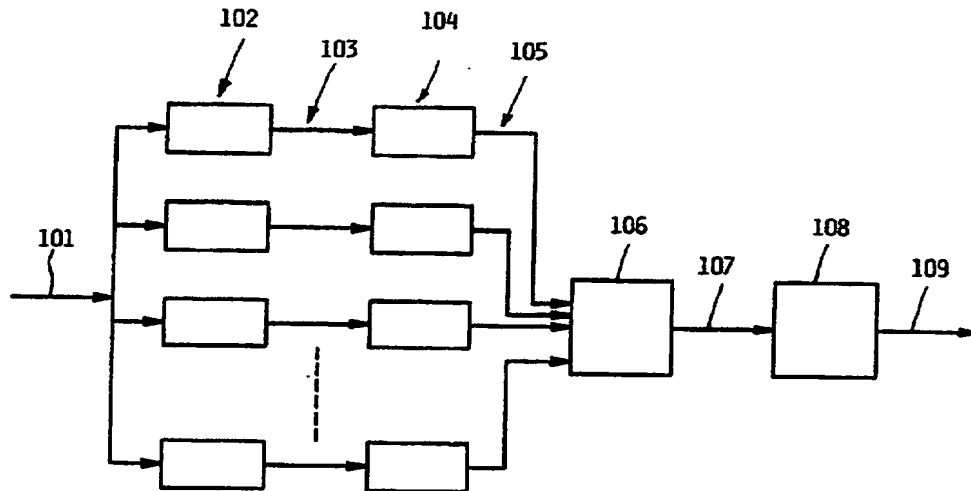
(57) Abstract

For classification of a mixed speech and noise signal (101) the signal is divided into separate, frequency limited subsignals (103), each of which contains at least two harmonic frequencies for the speech signal. The envelopes (105) of the subsignals (103) are formed as well as a measure (107) of synchronism between the envelopes (105). The synchronism measure (107) is compared with a threshold value for classification of the mixed signal as being significantly or insignificantly affected by the speech signal. The classification takes place with an unpresidented frequency and can therefore form the basis for a considerably more precise estimate of the noise signal than before, in particular when this has a speech-like nature.

* See back of page

DESIGNATIONS OF "DE"

Until further notice, any designation of "DE" in any international application whose international filing date is prior to October 3, 1990, shall have effect in the territory of the Federal Republic of Germany with the exception of the territory of the former German Democratic Republic.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	ES	Spain	MC	Monaco
AU	Australia	FI	Finland	MG	Madagascar
BB	Barbados	FR	France	ML	Mali
BE	Belgium	GA	Gabon	MR	Mauritania
BF	Burkina Fasso	GB	United Kingdom	MW	Malawi
BG	Bulgaria	GR	Greece	NL	Netherlands
BJ	Benin	HU	Hungary	NO	Norway
BR	Brazil	IT	Italy	PL	Poland
CA	Canada	JP	Japan	RO	Romania
CF	Central African Republic	KP	Democratic People's Republic	SD	Sudan
CG	Congo		of Korea	SE	Sweden
CH	Switzerland	KR	Republic of Korea	SN	Senegal
CM	Cameroon	LI	Liechtenstein	SU	Soviet Union
DE	Germany	LK	Sri Lanka	TD	Chad
DK	Denmark	LU	Luxembourg	TG	Togo
				US	United States of America

- 1 -

A method and an apparatus for classification of a mixed
speech and noise signal
----

5   The invention concerns a method and an apparatus for
classification of a mixed speech and noise signal as being
significantly or insignificantly affected by the speech
signal.

10   The time intervals where the mixed signal is insignifi-
cantly affected by the speech signal may be used for
forming a running estimate of the noise signal with known
methods, it being possible to suppress the noise on the
basis of this estimate.

15

The invention may be used in electroacustic systems for
transmission and signal processing of speech signals (e.g.
mobile telephones, speech recognition systems and hearing
aids), where it is endeavoured to eliminate or reduce de-
20   gradation of speech quality, speech recognition and speech
perception because of present background noise using noise
suppressing and/or speech enhancing methods.

Electroacustic systems for transmission and signal pro-
25   cessing of speech signals exist in numerous types and for
many different purposes. The expansive development in the
field of digital electronics, including particularly the
digital signal processors, has made it possible to employ
a plurality of methods not practically useful before in
30   connection with removing or suppressing, in real time, the
background noise, which occurs either acoustically simul-
taneously with the speech signal (e.g. in a helicopter
cockpit where machine and rotor noise affects the acoustic
communication from the pilot) or as an electric signal,
35   equivalent therewith, in the transmission system itself.

- 2 -

Such methods are known from the literature and are called
noise suppression or speech enhancement methods. Of these
methods may be mentioned adaptive filtering and spectral
subtraction. See e.g. (1) and (7). The aim of improving

5   the signal/noise ratio (the ratio of speech signal magni-
tude to noise magnitude) is that the methods are to
counteract the degradation of the reception caused by the
noise and the intelligibility of the transmitted speech
signal. Several of the known methods are based on a run-

10   ning estimate of the statistic characteristics of the
background noise, e.g. intensity and frequency content.
With a speech or pause detector time segments are identi-
fied with and without speech signal, respectively, and in
the segments exclusively containing background noise

15   (speech pauses) the characteristics of the noise may be
estimated by suitable signal analysis. Assuming a certain
stationarity of the background noise this estimate may be
used for adjusting the noise suppression or speech en-
hancement method until the next time the noise can be

20   estimated.

Several methods are described in the literature for dis-
tinguishing between voiced speech, unvoiced speech, and
pauses, both without and with background noise. See e.g.

25   (4), (5) and (8). (9) includes i.a. a survey of the most
important methods which have been used for classification
of speech, in particular in connection with speech recog-
nition systems.

30   In particular two of the known principles should be men-
tioned: the energy histogram and valley detector prin-
ciples. In a noise suppression method (3) use of the
valley detector method is reported for pointing out the
time intervals in which a mixed speech and noise signal

35   exclusively consists of background noise (i.e.
corresponding to pauses in the speech signal). In the

described invention the method is incorporated in a type of feedback loop by acting on the individual frequency bands of the output signal and with the purpose of increasing the field of use of the speech/noise detector.

5

However, none of the known speech and pause detectors are particularly robust when the speech signal is subjected to e.g. considerable reverberation, or when the background noise is added in a poor signal/noise ratio (less than 0
10   dB) or has a speech-like nature, i.e. resembles the speech signal from one or more speakers. In these cases the detection will be less certain with known methods. It has been attempted to reduce this problem by using a priori knowledge about the speech and noise signals. It has thus
15   been utilized in (1) and (2) that the amplitude fluctuations in speech and noise are different in certain cases. When, however, the noise is speech-like, this difference will be marginal.

20   So far, no speech detector has been developed which can operate reliably both with a poor signal/noise ratio and with speech-like noise. The object of the present invention is therefore to provide a method and an apparatus where this problem is solved.

25

This object is achieved by the method stated in claim 1 and the apparatus stated in claim 8, involving detection of the time segments in a mixed speech and noise signal which are dominated by the speech signal. This is to be
30   understood in combination with well-known knowledge, which is described below, that a speech signal includes a plurality of time segments where the speech signal contributes only insignificantly to the mixed signal. Such segments are not just speech pauses (between words and sen-
35   tences, breathing), but in particular also very short intervals, typically within a word where the speech signal

- 4 -

assumes a value so that it just contributes insignifi-
cantly to the mixed signal. These segments are detected,
and it is possible·on the basis of this to update para-
meters for the background noise. This is done with unpre-
5   cedented frequency and can therefore form the basis for a
considerably more precise estimate of the background
noise.

In a speech signal the energy can assume relatively great
10   values in short time intervals, corresponding to some of
the voiced sounds (e.g. the open vowels) as well as some
of the consonants (the fricatives and the plosives).
Therefore, the signal/noise ratio will be relatively great
in time segments containing these speech sounds, and these
15   segments are thus particularly useful for detecting pre-
sence of speech in background noise. The reason why the
energy is great in the mentioned speech sounds is the
following:

20   1) A vowel may be described as a (quasi)periodic time
signal which in terms of frequency consists of a funda-
mental frequency and its harmonics, whereby the speech
energy simultaneously occurs in a larger frequency range.

25   2) A fricative and/or a plosive may be described as a
short, noise-like time signal where the energy simul-
taneously occurs in a wide frequency range.

In the preferred embodiment of the invention the frequency
30   range of the speech signal is suitably divided into a plu-
rality of frequency bands, and it thus applies that for
each of the two types of speech sounds the energy occurs
with a certain simultaneousness between the frequency
bands. Further, it is special to the vowels that since the
35   difference between two consecutive harmonic frequencies is
always equal to the fundamental frequency for the speech

- 5 -

signal, the envelope of a frequency restricted subsignal
containing two or more consecutive harmonic frequencies
will always be periodic and substantially synchronous with
the fundamental frequency, since the envelope represents a
5   beat signal with a frequency equal to the difference be-
tween the two harmonics, which is precisely equal to the
fundamental frequency. Since it is the same frequency,
viz. the fundamental frequency of the speech signal, for
all the subsignals which causes the beat signal which is
10   detected by envelopment, the envelopes of the subsignals
will substantially be synchonous or correlated with each
other.

In order that this envelope, which is periodic with the
15   fundamental frequency, can always be produced, it is ne-
cessary that each subsignal has a frequency band width
which always comprises at least two harmonic frequencies.
This is obtained with a band width of at least twice the
fundamental frequency. If the fundamental frequency is
20   e.g. 220 Hz, the band width must at least be 440 Hz.

It is well-known from the literature, see e.g. (3), to
examine a mixed speech and noise signal by division into
time intervals and by splitting into a number of sub-
25   signals by means of a filter bank consisting of bandpass
filters. However, in contrast to the previously described
methods, this is done in a particular manner in the pre-
sent invention, since the invention realizes a filter bank
consisting of bandpass filters with a band width which is
30   especially dependant upon general characteristics of the
speech signal, as well as a detector utilizing the corre-
lation between the envelopes of the subsignals. Moreover,
and still in contrast to the previously described methods,
the aim of the present invention is not to point out the
35   time intervals in the mixed speech and noise signal which
just consist of noise (i.e. corresponding to pauses in the

- 6 -

speech signal), but to point out the intervals which are <u>dominated</u> by the speech signal.

The invention will be explained more fully by the follow-
5   ing description of a preferred embodiment with reference
to the drawing, in which

fig. 1 is a block diagram schematically showing an appa-
ratus according to the invention,
10
fig. 2 shows an example of an input signal consisting of a
portion of a speech signal without noise, and how this
signal is processed in the apparatus in fig. 1,

15   fig. 2A shows the input signal,

fig. 2B shows the frequency limited subsignals originating
from filtering of the input signal,

20   fig. 2C shows the envelope signals corresponding to the
subsignals in fig. 2B,

fig. 2D shows the synchonism signal from the synchronism
detector as well as a threshold value with which it is
25   compared, and

fig. 2E shows the final classification signal from the
threshold detector.

30   In fig. 1 an electric input signal 101 consisting of a
speech signal mixed with a noise signal (trafic noise,
cafeteria noise, speech from other persons or the like) is
passed to a filter bank 102 consisting of a plurality of
optionally overlapping bandpass filters with increasing
35   center frequency and covering in combination the entire
frequency range of the speech signal or part thereof. Each

bandpass filter has a band width greater than twice the
greatest expected value of the fundamental frequency of
the speech signal, so that a subsignal 103 comprising at
least two consecutive harmonic frequencies to the funda-
5   mental frequency can pass through each bandpass filter.

The subsignals are passed to their respective envelope
detectors 104, which form the time envelopes 105 for the
subsignals 103 e.g. by means of rectification, squaring or
10   analytical signals as well as optional subsequent low-pass
filtering. This signal processing, which following band-
pass filtering of the input signal generates and utilizes
the envelopes of the bandpass filtered subsignals is known
in other connections from the acoustic/audiological field,
15   see e.g. (6).

The envelope signals are passed to a synchronism detector
106, which produces a measure of synchronism between the
envelope signals 105 for a time segment of the signals.
20   Then, the time course of the computed synchronism has the
shape of a staircase curve and is called the synchronism
signal 107.

The principle of the synchronism detector 106 may e.g. be
25   based on correlation, an artificial neural network or
another computing method applied to all or a subset of the
envelope signals 105. For example, a correlation can be
computed by first computing the product sum of the signal
values for any pair of signals i.e. the envelope signals
30   from two adjacent bandpass filters and then performing
summation of all the computed product sums.

Finally, the synchronism signal 107 is passed to a thres-
hold detector 108 where the synchronism signal 107 is com-
35   pared with a threshold value. If the synchronism signal
107 is greater than the threshold value, the time segment

- 8 -

in question is classified as being dominated by speech,
and the classification signal 109 is set to the value
binary 1. If not, the classification signal 109 is set
to the value binary 0.

5

The overall function of the synchronism detector 106 and
the threshold detector 108 may also be implemented by
means of either a trained, a self-organizing or other
artificial neural network using the envelope signals 105
10   as input signals and forming the desired classification
signal 109 as output signal for classification of the
mixed signal.

Presence of a noise signal affects the classification more
15   or less depending upon the characteristics of the noise
signal. If the noise signal is stochastic, speech-like
noise, the speech detection will by and large not be af-
fected even with a very small signal/noise ratio. If, on
the other hand, the noise signal is a signal with an in-
20   herent modulation as a speech signal, or if it is a real
speech signal from one or more persons, the interplay be-
tween the actual signal/noise ratio and the construction
of the threshold detector 108 will be of decisive impor-
tance. When e.g. the threshold detector 108 is arranged
25   such that the threshold value 210 with a given time con-
stant adaptively adjusts itself corresponding to a given
fraction of the size of the synchronism signal 107, then
only the dominating speech signal will advantageously be
detected. Removal of the lowest frequency components of
30   the synchronism signal provides the additional advantage
that a continuous noise signal consisting of harmonic fre-
quency components (e.g. acoustic noise from a rotating
machine), will not erroneously be classified as being a
speech signal.

35

Fig. 2 shows an example of how a given input signal 201 is processed in the apparatus in fig. 1. To illustrate the fundamental principle of the invention the input signal 201 is shown in fig. 2A as a short speech signal without
5   noise consisting first of a (voiced) vowel and then of an unvoiced fricative. Fig. 2B shows the frequency limited subsignals 203 formed in the filter bank 102. Fig. 2C illustrates the envelope signals 205 formed by the envelope detectors 104 from the subsignals 203 in fig. 2B. At
10   the vowel, the envelope signals 205 in several frequency bands are shown to be correlated with each other and modulated with a frequency corresponding to the fundamental frequency. At the fricative, the envelope signals 205 show that short-term energy is present simultaneously in
15   several frequency bands. Fig. 2D shows the synchronism signal 207 computed from the synchronism detector 106 as well as the threshold value 210 with which it is compared. Finally, fig. 2E shows the obtained classification signal 209.

20
An apparatus according to the invention may be implemented either in analog or digital hardware or in software or in combinations thereof.

25

30

35

- 10 -

<u>References</u>:

(1)   US Patent No. 4 025 721

(2)   US Patent No. 4 185 168

(3)   US Patent No. 4 630 304

(4)   Cox B.V. and Timothy L.M.K. 1980. Nonparametric Rank-
      Order Statistics Applied to Robust Voiced-Unvoiced-
      Silence Classification. IEEE Trans. ASSP 28,5,550-
      561.

(5)   Gordos G. 1983. SPEECH DETECTION IN SEVERE NOISE.
      Proc. 11 ICA 91-94.

(6)   Houtgast T. and Steeneken H.J.M. 1973. The modulation
      transfer function in room acoustics as a predictor of
      speech intelligibility. Acoustica, 28, 66-73.

(7)   Lim J.S. 1986. SPEECH ENHANCEMENT. Proc. ICASSP 3135-
      3142.

(8)   McAulay R.J. and Malpass M.L. 1980. Speech Enhance-
      ment Using Soft-Decision Noise Suppression Filter.
      IEEE Trans. ASSP 28,2,137-145.

(9)   Savoji M.H. 1989. A robust algorithm for accurate
      endpointing of speech signals. Speech Comm. 8, 45-60.

- 11 -

P a t e n t   C l a i m s :
----

1. A method of classifying, in a selected time interval,
a mixed speech and noise signal (101, 201) as being signi-
ficantly or insignificantly affected by the speech signal,
where the mixed signal is divided into a plurality of se-
parate, frequency limited subsignals (103, 203), c h a -
r a c t e r i z e d in that

- each subsignal (103, 203) comprises at least two harmo-
nic frequencies for a fundamental frequency of the speech
signal,

- the time envelope (105, 205) is generated for the sub-
signals (103, 203),

- a measure (107, 207) of synchronism between these enve-
lopes (105, 205) is generated, and

- this measure (107, 207) is compared with a threshold
value (210).

2. A method according to claim 1, c h a r a c t e r -
i z e d in that the mixed signal is divided into a plu-
rality of time intervals in which the signal is classified
successively.

3. A method according to claim 1, c h a r a c t e r -
i z e d in that the selected time interval is a running
time window.

4. A method according to claims 1-3, c h a r a c t e r -
i z e d in that all envelopes are used for generating the
measure (107, 207) of synchronism between the envelopes
(105, 205).

- 12 -

5. A method according to claims 1-3, c h a r a c t e r -
i z e d in that one or more subsets of the envelopes
(105, 205) are used for generating the measure (107, 207)
of synchronism between the envelopes (105, 205).

5

6. A method according to claims 1-5, c h a r a c t e r -
i z e d in that the generation of the measure (107, 207)
of synchronism between the envelopes (105, 205) is based
on a correlation computation.

10

7. A method according to claims 1-5, c h a r a c t e r -
i z e d in that the envelopes (105, 205) are passed as
input signals to an artificial neural network which clas-
sifies the signal.

15

8. An apparatus for classification of a mixed speech and
noise signal (101, 201), comprising filter means each of
which permits passage of a subsignal (103, 203), c h a -
r a c t e r i z e d in that

20

- each subsignal (103, 203) contains at least two harmo-
nic frequencies for a fundamental frequency for the speech
signal, and that the apparatus moreover comprises

25   - means (194) for generating the time envelopes (105,
205) of the subsignals,

- means (106) for generating a measure (107, 207) of
synchronism between these envelopes, as well as

30

- means (108) for comparing the synchronism signal (107,
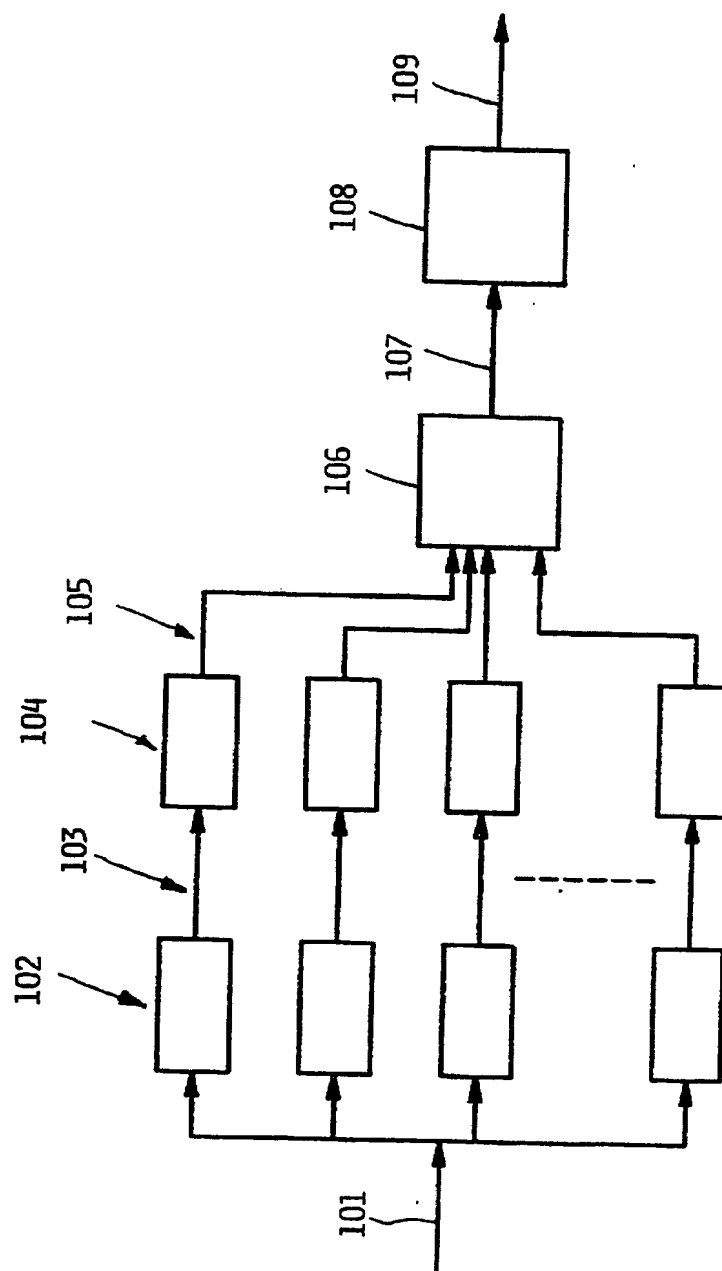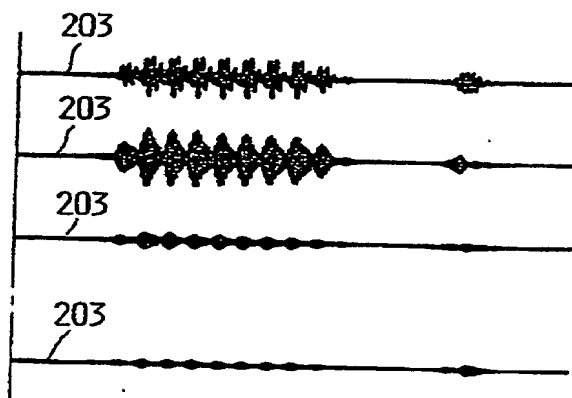207) with a given threshold value (210).

35

Fig. 1

2/2

Fig. 2A

201

time

Fig. 2B

203
203
203
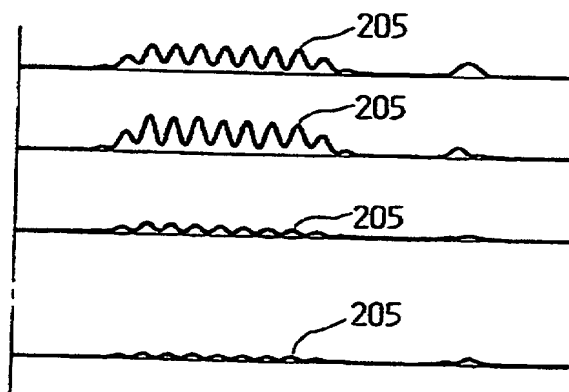203

Fig. 2C
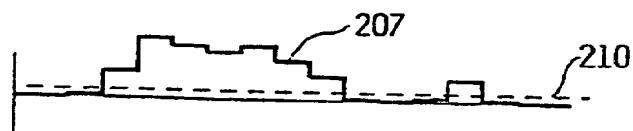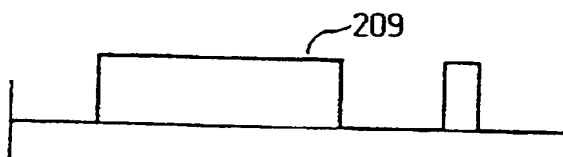

205
205
205
205

Fig. 2D

207
210

Fig. 2E

209

INTERNATIONAL SEARCH REPORT

International Application No PCT/DK90/00214

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all) *

According to International Patent Classification (IPC) or to both National Classification and IPC

IPC5: G 10 L 9/00

II. FIELDS SEARCHED

Minimum Documentation Searched 7

Classification System	Classification Symbols
IPC 5	G 01 L

Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched 8

SE, DK, FI, NO classes as above

III. DOCUMENTS CONSIDERED TO BE RELEVANT 9

Category *	Citation of Document, 11 with indication, where appropriate, of the relevant passages 12	Relevant to Claim No. 13
A	US, A, 4696039 (DODDINGTON) 22 September 1987, see the whole document	1-8
A	US, A, 4630304 (BORTH ET AL) 16 December 1986, see the whole document	1-8
A	US, A, 4382164 (MAY, JR.) 7 July 1983, see the whole document	1-8
A	US, A, 4277645 (MAY, JR.) 7 July 1981, see the whole document	1-8
A	DE, C2, 2649259 (FELTEN & GUILLEAUME FERNMELDEANLAGEN GMBH) 9 June 1983, see the whole document	1-8

---

* Special categories of cited documents: 10

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report
11th December 1990	20 NOV 1990
International Searching Authority	Signature of Authorized Officer
SWEDISH PATENT OFFICE	Rune Bengtsson

Form PCT/ISA/210 (second sheet) (January 1985)

ANNEX TO THE INTERNATIONAL SEARCH REPORT
ON INTERNATIONAL PATENT APPLICATION NO.PCT/DK 90/00214

This annex lists the patent family members relating to the patent documents cited in the above-mentioned international search report.
The members are as contained in the Swedish Patent Office EDP file on   90-09-27
The Swedish Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
US-A- 4696039	87-09-22	DE-A-	3473373	88-09-15
		EP-A-B-	0140249	85-05-08
US-A- 4630304	86-12-16	EP-A-	0226613	87-07-01
		JP-T-	63500543	88-02-25
		WO-A-	87/00366	87-01-15
US-A- 4382164	83-05-03	CA-A-	1158174	83-12-06
		CA-A-	1164351	84-03-27
		DE-A-	3101775	82-01-07
		FR-A-B-	2478909	81-09-25
		FR-A-B-	2482389	81-11-13
		GB-A-B-	2068698	81-08-12
		GB-A-B-	2136253	84-09-12
		JP-A-	56142600	81-11-06
		NL-A-	8100323	81-08-17
		US-A-	4277645	81-07-07
US-A- 4277645	81-07-07	CA-A-	1158174	83-12-06
		CA-A-	1164351	84-03-27
		DE-A-	3101775	82-01-07
		FR-A-B-	2478909	81-09-25
		FR-A-B-	2482389	81-11-13
		GB-A-B-	2068698	81-08-12
		GB-A-B-	2136253	84-09-12
		JP-A-	56142600	81-11-06
		NL-A-	8100323	81-08-17
		US-A-	4382164	83-05-03
DE-C2- 2649259	83-06-09	NONE		